Date, No.

Gradient Descent

<Gradient Descent>

- How do I take a "loss function" and turn it into a search strategy? (That's where Gradient Descent comes in)

- Gradient Descent
  : Refers to the process of walking down the surface formed by using our loss function on all the points in parameter space.

  * we'll only know loss values at the points in parameter space where we've evaluated our loss function.

  (while the loss is greater than a tiny constant compute the direction.)

  And then for ~~search~~ each parameter in the model,
  Set its value to be the old value plus the product of the step size and the direction.
  Then <u>finally</u> re-compute the loss.

  The <u>slope</u> or the <u>rate</u> at which the curve.
  → gives us a descent sense of how far to step and the direct at the same time.

Date . . No.

<Troubleshooting a Loss Curve>

Generally, though,
learning rate is a fraction significantly less than one.

Remember this formulation of gradient descent, and that
learning rate is a hyperparameter that is fixed during training.